Conflow: An Online Concept-Mapping Tool to Measure Conceptual Fluency

Shivin AGARWAL[a +], **Anveshna SRIVASTAVA**[a] **& Nisheeth SRIVASTAVA**[a]
[a] *Indian Institute of Technology Kanpur, Kanpur, India*
\+ shivin.agarwal15@gmail.com; shivin23@iitk.ac.in

Abstract: Assessment at scale operates under constraints of reproducibility, interpretability, and comparability. In practice, these constraints confine pedagogical evaluations to recitations of fact or reproductions of formulaic calculations. Efforts to design novel, more holistic methods of assessment tend to violate one of the three contracts listed above and thus find limited uptake in practice. In this paper, we describe *ConFlow* - an online assessment instrument that builds on the strengths of concept mapping and aims to satisfy these constraints while enabling holistic, concept-level assessment. We describe the tool and a validation study we undertook to evaluate its performance vis-à-vis personalized ratings provided by college undergraduates. ConFlow lets learners build concept maps, analyze them, and present a 'quadrant graph' depicting the mapper's fluency with the mapped concepts. The quadrant graph categorizes conceptual fluency as 'Well-Understood', 'Partially Understood', 'Fuzzy', and 'Unknown' concepts. ConFlow generates this graph based on the number of conceptual moves and the accuracy of the conceptual links and phrases made during the online mapping process. In this pilot work, we present our attempt at validating the tool for wider usage by the learning community. Our process-based assessment method has implications for capturing learning trajectories, and it offers a more profound understanding of concept-level uncertainties and misconceptions.

Keywords: ConFlow; Conceptual fluency; Concept-mapping; Quadrant-graph; Assessment

1. Introduction

Learning is as diverse as the people who engage in it. For one student, something that comes naturally may prove difficult for another. Despite this diversity, educational assessments use standardized tests to assess everyone using the same scale, providing structure but overlooking the complexity of unique learners. Although evaluation approaches have advanced significantly, techniques for assessment at scale struggle to satisfy three fundamental criteria: interpretability, comparability, and reproducibility.

Interpretability suggests that instructional methods and assessment outcomes must be transparent and understandable, aiding in comprehending the underlying reasoning behind the use of educational interventions (Baird, et al. 2019). Comprehension is essential since it helps instructors recognize which teaching strategies work or do not work in different pedagogical contexts. The 'comparability' criterion is crucial for systematically assessing the efficacy of various interventions across different contexts. It calls for diverse educational practices to be measured against uniform standards and metrics (Evans & Lyons, 2017). Without comparability, differences in implementation or contextual factors can obscure the actual impact of a teaching method, making it difficult to generalize findings or conduct meta-analyses that inform policy and practice (American Educational Research Association [AERA], APA, & NCME, 2014). The last criterion of 'reproducibility' ensures that a technique or intervention will produce consistent results when used in similar circumstances (Schuwirth & Van der Vleuten, 2018). Consistency is essential for scaling effective practices across diverse educational environments as well as for validating prior knowledge and the success of the educational outcome. Reproducibility in education is especially difficult because classrooms are dynamic and eclectic, with differences in student backgrounds, teacher skills, curriculum design, and even daily classroom conditions impacting results.

In this study, we explore the online concept-mapping process to tackle the above assessment constraints of interpretability, comparability, and reproducibility, along with developing a holistic assessment of a learner's fluency with individual concepts, based on both their learning process and outcome. Almost half a century ago, David Ausubel suggested that the most important factor influencing learning is what the learner already knows and that teachers should ascertain their prior knowledge and teach accordingly (Ausubel, 2012). Since the construction of concept maps involves the invocation of prior knowledge, along with the use of different modalities of text and visuals in establishing links across concepts, concept maps could potentially be used to test Ausubel's hypothesis. Hence, in this study, we set out to introduce and validate 'ConFlow', an online concept-mapping tool to assess learners' fluency with different concepts.

1.1 Concept Maps

Concept mapping is a graphical representation of a person's knowledge of a domain, where nodes (concepts), linking arrows, and phrases are used to provide insights into their mental organization of concepts (Novak, 2012). Nodes, connected by linking arrows and phrases, constitute 'propositions,' which are evaluated, along with the overall map structure, to assess conceptual understanding (Novak & Gowin, 1984). They can effectively address the challenges of interpretability, comparability, and reproducibility in education. They enhance interpretability as they present information visually in a structured node-link format, enabling teachers and students to quickly identify gaps in understanding and integrate new information with existing knowledge (Novak, 2012). They support comparability as their standardized formats, when evaluated with established rubrics, allow educators to objectively assess and benchmark the depth and clarity of students' cognitive structures (Katagall, Dadde, Goudar, & Rao, 2015). Lastly, as they require learners to externalize their internal knowledge frameworks, concept maps offer a reproducible artifact that can be consistently generated and reviewed over time.

Concept maps face several limitations, such as building simplistic maps that do not capture the complexity of the domain, a lack of metacognitive engagement that impedes learners' ability to structure and internalize new information (Mayer & Moreno, 2002), a lack of objectivity in assessment that impedes comparability across contexts (Novak, 2012), the static nature of maps limiting the fostering of deep understanding (Machado & Carvalho, 2020), resistance from students when it is not meaningfully integrated with the classroom routine (Maker & Zimmerman, 2020), and difficulty with revision and reorganization when using the pen-paper mode. As education systems become increasingly data-driven and digital, the opportunity for creating a dynamic concept mapping tool that performs a holistic assessment by letting learners metacognitively engage with their learning process is ripe. Hence, here, we propose an approach that leverages advances in adapting concept mapping toward pedagogical evaluations (Srivastava et al., 2021). In an earlier work, we had shown how procedural analysis could be used to measure learners' conceptual fluency (Srivastava et al., 2014). In the present work, we have moved to an online interface, designing *ConFlow*, that lets learners build concept maps, analyze them, and present a quadrant graph that depicts the mapper's fluency with the different mapped concepts.

2. Method

2.1 The Tool: ConFlow

The 'ConFlow' application allows users to register and choose their preferred domain. Currently, it hosts concepts from basic statistics and biology. For the present validation task, our focus was the 3-D structure of the DNA (Deoxyribonucleic Acid) biomolecule.

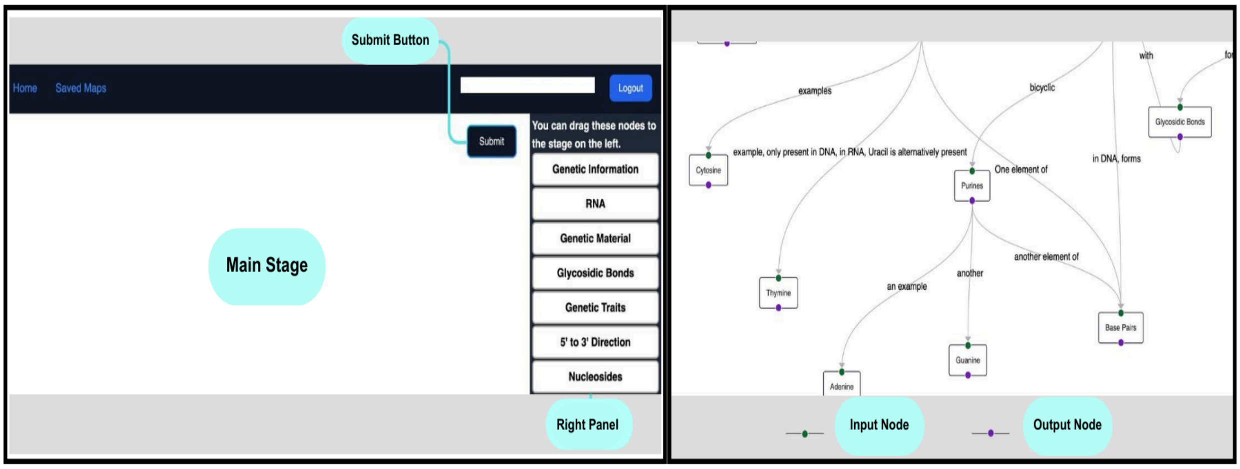

Figure 1. a) The Interface Of ConFlow, and b) A sample concept map

When the user reaches the mapping interface, they can see a list of draggable concepts on the right (see Figure 1a). To create a concept map, users can individually drag and drop these concepts onto the main stage. Two concepts can be connected with a linking arrow, indicating the directionality of the relationship, and with a linking phrase, suggesting the relationship between the two concepts. Thus, a map is built after all the concepts have been exhausted and linked to the conceptual network on the stage (see Figure 1b for a sample).

2.2 Categorization of Quadrants

We use two measures to characterize students' conceptual fluency. The number of moves per concept, plotted on the 'X'-axis, which we use as a proxy for the confidence that a user exhibits regarding the conceptual fluency with a concept, and the accuracy score based on their final map evaluation, plotted on the 'Y'-axis. Based on the above two measures, we classify the conceptual fluency of a learner with all the domain concepts used in the study into four quadrants: 'Unknown', 'Fuzzy', 'Partially Understood', and 'Well Understood' (see Figure 2). Unknown concepts have fewer moves and less accuracy, indicating incorrect connections and incorrect linking phrases. Fuzzy concepts have higher moves and less accuracy, indicating slightly correct connections to sibling concepts and correct linking phrases. Partially Understood concepts have more moves than the average and a higher final score, indicating correct connections and correct linking phrases. Well Understood concepts have fewer moves and more accuracy, indicating correct connections and correctly marked linking phrases.

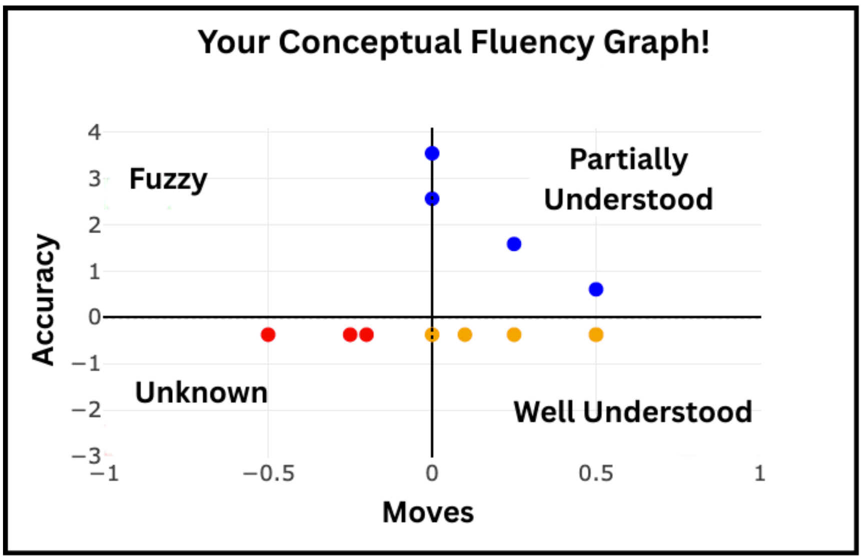

Figure 2. The quadrant graph: dots signify concepts, and the categories signify fluency

Evaluation of the concept maps involved assigning '0' to no linking phrase, '0.5' to incomplete sentences, '0' to meaningless propositions, and '1' to correct propositions.

2.3 Sample and the task

The group of participants in the validation study consisted of 16 consenting adults (Males=10, mean age=21.6 years, SD=2.98; Females=6, mean age=23.3 years, SD=3.01). The participants viewed a 5-minute introductory video on how to use the tool. A training task involving a sample concept map was provided to acclimatize them with the interface. The main task then commenced, where each participant worked with 26 concepts to build the concept map on DNA. Each participant's comprehension of the mapping was assessed through a personal rating after task completion. After the task, participants were asked to classify their understanding of individual concepts into one of four predefined categories: 1. Unknown: You don't know about the concept, 2. Fuzzy: You have a vague understanding of this concept, having encountered it before, but you don't recall much about it, 3. Partially Understood: You know the concept but can't recall it completely, 4. Well Understood: You are very confident about your knowledge of the concept. After the main task, participants were administered a post-test in a multiple-choice questionnaire format to assess their understanding of concepts related to the building task. Lastly, they were asked to complete a feedback form to share their overall experience. The entire experiment was completed within 60 minutes.

3. Analysis

3.1 Evaluation of Concept Maps in ConFlow

We use Cohen's Kappa (κ), a statistical measure, to quantify the level of agreement between the two ratings of conceptual fluency- personal and the tool-based conceptual categorizations. Cohen's Kappa is calculated as:

$$K = (Po - Pe)/(1-Pe)$$

We operationalize 'Po' as the ratio of the common concepts in that category to the total number of concepts marked in that category. 'Pe' is the probability that suggests that a single concept can come in any of the four categories. Here, it is 1/4 (probability of random assignment to any of the four quadrants).

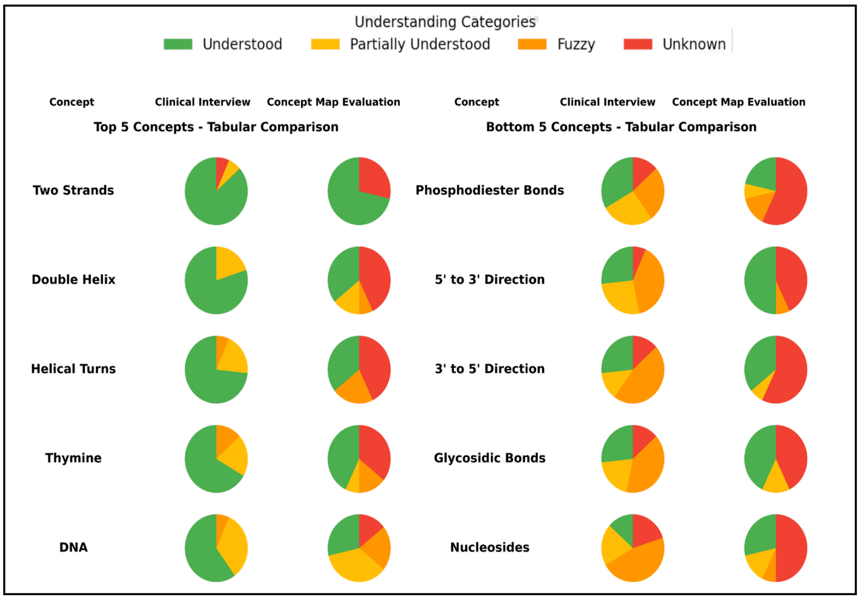

Figure 3. The best and the least well-understood concepts in the study

4. Results

Our validation strategy involved estimating the extent to which our novel concept-mapping-based measurement of gradation in conceptual fluency could reproduce similar gradation inferred via personal ratings. These estimates in turn were obtained from the quadrant plots defined above in the Methods section, with each subject's quadrant plot used to identify whether a concept was well understood, partially understood, fuzzy, or unknown. The pie charts in Figure 3 show the relative fraction of students falling into each category for our measurement instruments. Visually, it is evident that the concept mapping results align well with the clinical interview-based judgment for the study population. The Spearman rank correlation of the degree of conceptual understanding, measured across all 26 concepts using both these methods, is 0.225, further supporting the observation that the concept map measurement is a weak to moderate proxy for personal rating at a cohort level.

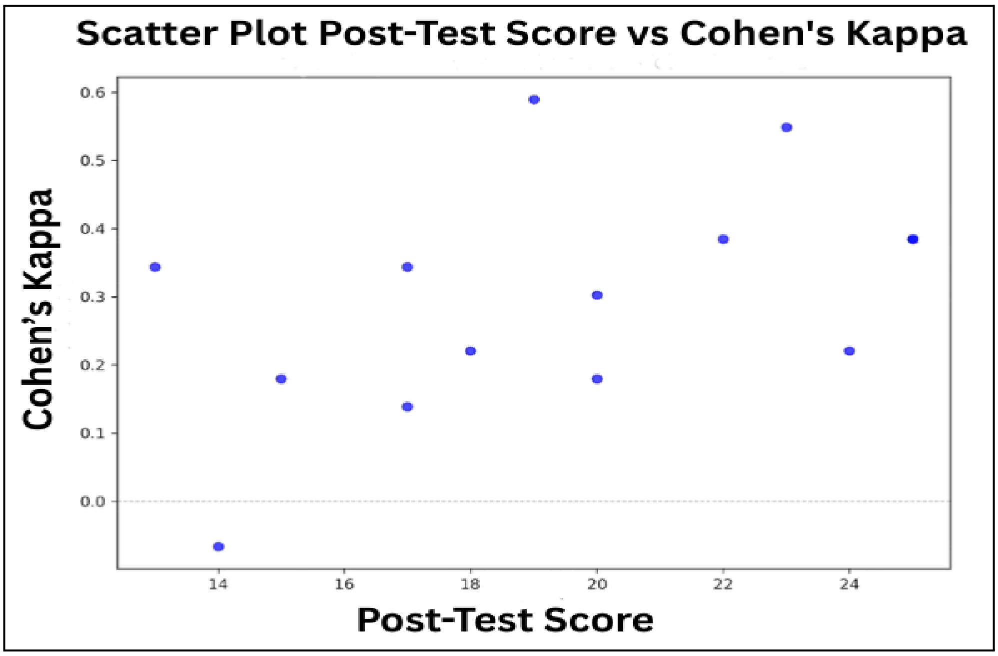

Figure 4. Post Test vs Cohen's Kappa relation

Interestingly, this relationship is stronger at the individual level. Figure 4 plots the individual-level proportion of agreement (measured using Cohen's kappa) for all participants on the Y-axis, with the average (mean = 0.297) significantly greater than zero (p-value: < 0.001). Cohen's kappa calculation uses understanding codes assigned to a candidate based on personal ratings of concepts and codes assigned to a candidate based on the quadrant graph to estimate a within-participant proportion of agreement for the two modalities. We also find that performance in the study post-test highly correlates with the proportion of agreement between the two measurement modalities ($r = 0.615$). We interpret this correlation to reflect variability in effort level in the study procedure by students since the post-test was intended to be substantially straightforward and thus was difficult to score poorly on without inattentive performance.

5. Discussion, Limitations, & Way Forward

In this work, we examined how concept maps could enhance pedagogical evaluations by producing reproducible, comparable, and interpretable measurements that are significantly more holistic and granular than traditional methods. *ConFlow* demonstrates a way to We measure granular categories, such as "well understood" and "fuzzy," to bridge the gap between qualitative insights and quantitative reproducibility. The digital interface represents a novel assessment approach, allowing broader application and scaling-up possibilities across domains. The results indicate that individuals have a poorer sense of what they do not know (Figure 3), which could be utilized in future studies to enhance learners' metacognitive abilities. Owing to the limited dataset, generalization is complex. However, this pilot study contributes meaningfully to the literature on process analysis, focusing on learning processes rather than solely on outcomes (Zhang, Wang, Qi, Liu, & Ying, 2023).

This is in line with the position that persistence and effort shown in a domain reflect the development of resilience and grit, which are necessary for achieving a more profound understanding (Dweck, 2006). In our work, we expect concepts to gradually move from the 'unknown' zone to the 'known' zone, based on learners' fluency with those concepts. This behavior would demonstrate the growing competence of learners in the domain. Just as sustained effort over time predicts success more reliably (Duckworth, 2016), we hypothesize that the movement of concepts across quadrants would determine learners' persistence and fluency as well. Taking notes from reviewers' insights, future work will involve empirical testing in real classrooms and modifying the dashboard to include instructional next steps and feedback loops. We also intend to undertake a longitudinal study with a larger participant sample to further validate the system's effectiveness.

Acknowledgments
We are grateful to all the participants for their contribution to this study.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). Standards for educational and psychological testing. Washington, DC: American Education Research Association and National Academy of Education.

Ausubel, D. P. (2012). The acquisition and retention of knowledge: A cognitive view. Springer Science & Business Media.

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. Educational Researcher, 48(4), 217- 228.

Dweck, C. S. (2006). Mindset: The new psychology of success. Random House.

Duckworth, A. (2016). Grit: The power of passion and perseverance (Vol. 234). New York, NY: Scribner.

Evans, C. M., & Lyons, S. (2017). Comparability in balanced assessment systems for state accountability. Educational measurement: issues and practice, 36(3), 24-34.

Katagall, R., Dadde, R., Goudar, R. H., & Rao, S. (2015). Concept mapping in education and semantic knowledge representation: an illustrative survey. Procedia computer science, 48, 638-643.

Novak, J. D. (2012). Concept mapping: A strategy for organizing knowledge. In Learning science in the schools (pp. 229-245). Routledge.

Novak, J. D., & Gowin, D.B. (1984). Learning how to learn. Press Syndicate of the University of Cambridge.

Schuwirth, L. W., & Van der Vleuten, C. P. (2018). How to design a useful test: the principles of assessment. Understanding medical education: evidence, theory, and practice, 275-289.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. Learning and instruction, 12(1), 107-119.

Machado, C. T., & Carvalho, A. A. (2020). Concept mapping: Benefits and challenges in higher education. The Journal of Continuing Higher Education, 68(1), 38-53.

Maker, C. J., & Zimmerman, R. H. (2020). Concept maps as assessments of expertise: Understanding of the complexity and interrelationships of concepts in science. Journal of Advanced Academics, 31(3), 254-297.

Srivastava, A., Srivastava, N., & Chandrasekharan, S. (2014). Measure concept-mapping, not concept-maps: Procedural analysis elucidates stages in students' understanding of biology concepts. Proceedings of the annual meeting of the American Educational Research Association.

Srivastava, A., Srivastava, N., & Chandrasekharan, S. (2021). Quality of concept maps is affected by map-building strategies. Biochemistry and Molecular Biology Education, 49(1), 129-139.

Zhang S, Wang Z, Qi J, Liu J, Ying Z. Accurate Assessment via Process Data. Psychometrika. 2023;88(1):76-97. doi:10.1007/s11336-022-09880-8